AFRL-AFOSR-JP-TR-2016-0039

Dynamic Trust Models between Users over Social Networks

Kazumi Saito
University Of Shizuoka

04/05/2016
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ IOA
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
23-06-2016	Final	31-03-2013 to 30-03-2016

4. TITLE AND SUBTITLE
(134042) Dynamic Trust Models between Users over Social Networks

5a. CONTRACT NUMBER
FA2386-13-1-4042

5b. GRANT NUMBER
Grant 13RSZ066_134042

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Prof Kazumi Saito

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
School of Administration and Informatics, University of Shizuoka
52-1 Yada, Suruga-ku
Shizuoka 422-8526
Japan

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AOARD
UNIT 45002
APO AP 96338-5002

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR/IOA(AOARD)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AOARD-134042

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution A: Approved for public release. Distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
In this project, by focusing on a number of word-of-mouth communication websites, we attempted to construct dynamic trust models between users that enable to explain trust formation and its evolution processes over social networks with reasonable accuracy. Consequently, we obtained our research results on four new approaches, 1) to rank items based on the MTDF (Multinomial with Trust Discount Factor) model, 2) to estimate the conformity of users from the observed review scores, 3) to predict evolution of trust links under the presence of mediators, and 4) to analyze activities among users based on a non-negative matrix factorization (NMF) method. By using the datasets of item review scores and trust networks collected from a number of websites such as Epinions, we confirmed that our proposed models and methods can be useful basic techniques for uncovering fundamental mechanisms of trust formation and its evolution processes over trust networks. During our research period, we presented 3 papers published in peer-reviewed journals and 13 papers published in peer-reviewed conference proceedings as attached in this report.

15. SUBJECT TERMS
Social Network, Computer Science, Data Mining, Trust

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** / **19b. TELEPHONE NUMBER** *(Include area code)*
U	U	U	SAR	12	Hiroshi Motoda, Ph. D. / +81-3-5410-4409

Final Report for AOARD Grant 134042

"Dynamic Trust Models between Users over Social Networks"

30/03/2016

Name of Principal Investigators (PI and Co-PIs): Kazumi Saito

- e-mail address : k-saito@u-shizuoka-ken.ac.jp
- Institution : School of Administration and Informatics, University of Shizuoka
- Mailing Address : 52-1 Yada, Suruga-ku, Shizuoka 422-8526 Japan
- Phone : +81-54-264-5436
- Fax : +81-54-264-5436

Period of Performance: 03/31/2013– 03/30/2016

Abstract: Short summary of most important research results that explain why the work was done, what was accomplished, and how it pushed scientific frontiers or advanced the field. This summary will be used for archival purposes and will be added to a searchable DoD database.

We obtained our main research results on four new approaches, 1) to rank items based on the MTDF (Multinomial with Trust Discount Factor) model [C9], 2) to estimate the conformity of users from the observed review scores [C4], 3) to predict evolution of trust links under the presence of mediators [C5], and 4) to analyze activities among users based on a non-negative matrix factorization (NMF) method [C1].

First, we proposed a new item-ranking method that is reliable and can efficiently identify high-quality items from among a set of items in a given category using their review-scores which were rated and posted by users [C9]. Typical ranking methods rely only on either the number of reviews or the average review score. Some of them discount outdated ratings by using a temporal-decay function to make a fair comparison between old and new items. The proposed method reflects trust levels by incorporating a trust discount factor into a temporal-decay function. More specifically, we first defined the MTDF (Multinomial with Trust Discount Factor) model for the review-score distribution of each item built from the observed review data. We then brought in the notion of z-score to accommodate the trust variance that comes from the number of reviews available, and proposed a z-score version of MTDF model. Finally we demonstrated the effectiveness of the proposed method using the MovieLens dataset, showing that the proposed ranking method can derive more reasonable and trustable rankings, compared to two naïve ranking methods and the pure z-score based ranking method.

Second, we proposed a simple and efficient method that learns and assesses the conformity of each user of an online review system from the observed review score record [C4]. The model we use is a modified Voter model that takes account of the conformity of each user. Conformity is learnable quite efficiently with a few tens of iterations by maximizing the log-likelihood given the observed data. The proposed method was evaluated and confirmed effective by two review datasets. It could identify both high and low conformity users. Users with high conformity were not necessarily early adopters. Their scores are influential to drive the consensus score. The user ranking of conformity was compared with PageRank and HITS in which user network was roughly approximated by the directed graph induced by the observed data. The proposed method gave more interpretable ranking, and the global property of high conformity users was identified.

Third, we analyzed evolution of trust networks in social media sites from a perspective of mediators. To this end, we proposed two stochastic models that simulate the dynamics of creating a trust link under the presence of mediators, the A-ME and A-MAE models, where the A-ME model analyzes mediator effects for trust-network evolution in terms of mediator types, and the A-MAE model, an extension of the A-ME model, analyzes mediator-activity effects for trust-network evolution [C5]. We presented an efficient method of inferring the values of model parameters from an observed sequence of trust links and user activities. Using real data from Epinions, we experimentally showed that the A-MAE model significantly outperforms the A-ME model for predicting trust links in the near future under the presence of mediators, and demonstrated the effectiveness of mediator-activity information for trust-network evolution. We further clarified, by using the A-ME and A-MAE models, several characteristic properties of trust-link creation probability in the Epinions data.

Finally, we analyzed evolution of activities among users for an item-review site based on non-negative matrix factorization (NMF) methods that have recently been shown useful for trust-link prediction in such a site where both link and activity information is available. Here, a user activity in an item-review site means posting a review and giving a rating for an item. Towards better trust-link prediction, we proposed a new NMF method that incorporates people's evaluation of users' activities as well as trust-links and users' activities themselves [C1]. We further applied it to an analysis of users' behavior. Using two real world item-review sites, @cosme and Epinions, we statistically analyzed the datasets, and in particular confirmed that the number of appreciation messages received correlates with the number of trust-links received, suggesting that incorporating the activity-evaluation information can be a promising approach. Next, we demonstrated that the proposed method outperforms the state-of-the-art hTrust and its variants for solving the trust-link prediction problem.

In addition to the above main research results, we developed a number of fundamental techniques for efficiently analyzing influential nodes in social networks [J1, J2, C8], effectively detecting changes in time-series data [J3, C2, C6], and reliably performing re-sampling simulations for large-scale networks [C3, C7, C10]. Here we should node that these techniques played important roles to obtain our main research results.

Introduction: Include a summary of specific aims of the research and describe the importance and ultimate goal of the work.

People, e.g., Internet users, constantly receive/send a large number of messages from/to other people, and various kinds of information diffuse over time. Through such human interactions, trust relations between users (or conformity of users) are formed and evolve over time. These kinds of trust formation and its evolution processes are mostly characterized by individual phenomena over social networks, which is quite complex, but there should be some regularities. It would be possible to find empirical regularities and develop explanatory accounts of these processes in terms of macroscopic statistical properties. Furthermore, by constructing computational models based on these statistical properties, we can expect to precisely estimate how much information diffuse and which opinions prevail in future. Especially, such predictive capability would be valuable for anticipating social trends, and market opportunities. Thus, we propose to conduct research on computational models and methods for uncovering fundamental mechanisms of trust formation and its evolution processes over social networks. In this project, by focusing on a number of word-of-mouth communication websites, we first attempt to construct dynamic trust models between users that enable to explain trust formation and its evolution processes over social networks with reasonable accuracy. Then, based on these fundamental models, we plan to develop more advanced models for information diffusion and opinion formation, together with several techniques for detecting users' characteristics behaviors.

Experiment: Description of the experiment(s)/theory and equipment or analyses.

Trust Discount Modeling [C9]: We denote the sets of users and items by $V = \{u, v, w, \dots\}$ and $I = \{i, \dots\}$, respectively. When a user $v \in V$ reviewed an item $i \in I$, we denote its timestamp and score by $t(v, i)$ and $s(v, i)$, respectively, where each score $s(v, i)$ is denoted by a positive integer in $S = \{1, \dots, |S|\}$, and $|S|$ stands for the number of elements in S. Then, we can express our observed data set as $D = \{\dots, (v, i, t(v,i), s(v,i)), \dots\}$. Hereafter, let $V(i) = \{v \mid (v, i, t(v,i), s(v,i)) \in D\}$ be a set of users who reviewed an item i. For users in $V(i)$, let $U(i, t) = \{u \in V(i) \mid t(u,i) < t\}$ be the set of users whose review times are before t, and $U(i, t, s) = \{u \in U(i, t) \mid s(u, i)= s\}$ the set of those users whose review score is s. In general, users may decide their review scores of each item by taking account not only of their own evaluations, but also of past majority scores or those submitted by high conformity users. In order to stochastically cope with the opinion decision problem affected by majority scores, we can employ the basic voter model, and define the probability that a user v gives a score s to an item i at time t as $P_0(s \mid i, t) = (1 + |U(i, t, s)|)/(|S| + |U(i, t)|)$, where we employed a Bayesian prior known as the Laplace smoothing. Here we note that the Laplace smoothing corresponds to the assumption that each node initially holds one of the $|S|$ scores with equal probability. Note also that the Laplace smoothing corresponds to a special case of Dirichlet distributions that are very often used as prior distributions in Bayesian statistics. We refer to this model as the base multinomial model.

As for the base multinomial model, we assumed that all the past reviews are equally weighted. However, it is naturally conceivable that some of the quite old reviews are almost out-of-date and their trust levels might be low. In order to reflect this kind of effects into the model, we consider introducing some trust discount factors. The simplest one is an exponential discount factor defined by $\rho(\Delta t; \lambda) = \exp(-\lambda\Delta t)$, where $\lambda \geq 0$ is a parameter and $\Delta t = t - t'$ stands for the time difference between t and t'. Another natural one would be a power-law discount factor defined by $\rho(\Delta t; \lambda) = (\Delta t)^{-\lambda} = \exp(-\lambda \log \Delta t)$, where $\lambda \geq 0$ is a parameter. Now, we construct a more general discount factor. For a given positive integer J, we consider a J-dimensional vector consisting of linearly independent features, $\mathbf{F}J(\Delta t) = (f1(\Delta t), \dots , fJ (\Delta t))^T$, and a parameter vector with nonnegative elements for these features, $\boldsymbol{\lambda}J = (\lambda 1, \dots , \lambda J)^T$. Then, we define a general discount factor by $\rho(\Delta t; \boldsymbol{\lambda}J) = \exp(-\boldsymbol{\lambda}J^T \mathbf{F}J(\Delta t))$. Using this general discount factor $\rho(\Delta t; \lambda)$, we define the MTDF (Multinomial with Trust Discount Factor) model in the following way. In our model, the base multinomial model is replaced with $P(s \mid i, t; \boldsymbol{\lambda}J) = (1 + \sum_{u\in U(i, t, s)} \rho(t - t(u,i); \boldsymbol{\lambda}J))/(|S| + \sum_{u\in U(i, t)} \rho(t - t(u,i); \boldsymbol{\lambda}J))$ for $k = 1, \dots, K$. Note that $P(s \mid i, t; \boldsymbol{\lambda}J)$ is reduced to $P_0(s \mid i, t)$ when $\boldsymbol{\lambda}J$ is the J-dimensional zero-vector $\mathbf{0}J$, that is, the MTDF model of $\boldsymbol{\lambda}J = \mathbf{0}J$ coincides with the base multinomial model. Here, we can estimate the trust discount parameter values $\boldsymbol{\lambda}J$ of the MTDF model by maximizing the likelihood function based on $P_0(s \mid i, t)$ for a given observed review results D. Note that the MTDF model of $\boldsymbol{\lambda}J \approx \mathbf{0}J$ for some item v means that this item does not need to introduce a trust discount factor, which maintains a high trust level. Therefore, we can construct a method of ranking items based on the MTDF model using the observed review results.

In our experiments on trust discount modeling, we employed the MovieLens 10M/100k dataset to experimentally evaluate our ranking methods. MovieLnes is one of the online movie recommender services, and the dataset consists of 10,000,054 ratings with time stamps that are made on a 5-star scale with half-star increments for 10,681 movies by 71,567 users. Assuming that it is drawn from a multinomial distribution with K = 10, the average score over all movies is 3.51 and the standard deviation is 1.06. Interestingly, the user is more likely to evaluate movies without using a half-star. Moreover, we can observe that many of the movies having over 10,000 reviews get relatively high scores greater than the overall average 3.51.

User Conformity Modeling [c4]: As for the base multinomial model, we assumed that all the past user scores are equally weighted independent of users. However, it is naturally conceivable that some high conformity users should have larger weights. In order to reflect this kind of effects into the model, we consider introducing a positive conformity metric $\exp(\theta(u))$ to each user u, where $\theta(u)$ is a parameter. Hereafter, we denote the vector consisting of these parameters by $\boldsymbol{\theta} = (\ldots, \theta(u), \ldots)$. Then, we can extend the base multinomial model $P_0(s \mid i, t)$ and build a generative model in which user v gives score s for item i at time t with the following probability. $P(s \mid i, t; \boldsymbol{\theta}) = (1 + \sum_{u \in U(i,t,s)} \exp(\theta(u)))/(|S| + \sum_{u \in U(i,t)} \exp(\theta(u))$. In our study, in order to estimate $\boldsymbol{\theta}$ from the observed data set D, we consider maximizing the following logarithmic likelihood function based on $P(s \mid i, t; \boldsymbol{\theta})$.

In our experiments on user conformity modeling, we collected review score records from two famous review sites in Japan and constructed two datasets for this experiment. One consists of review scores for cosmetics extracted from "@cosme" which is a Japanese word-of-mouth communication site for cosmetics. We refer to this dataset as the Cosmetics review dataset. The other one is composed of review records collected from "anikore", a ranking and review site for anime, which is referred to as the Anime review dataset. In both the datasets, each record has 4-triple (u, i, s, t), which means user u gives a score s to item i at time t. The Cosmetics review dataset has 297, 453 review records by 10, 403 users for 46, 398 items from 2008/12/07 to 2009/12/09, while the Anime review dataset has 300, 327 records by 13, 112 users for 1, 790 items from 2010/8/01 to 2012/8/08. Thus, the average numbers of reviews per user and item were 28.6 and 6.4 in the Cosmetics dataset, and 22.9 and 167.8 in the Anime dataset. The score is an integer value ranging from 1 to 7 and its average of overall ratings was 4.4 in the Cosmetics dataset, and from 1 to 5 and 3.9 in the Anime dataset.

Trust Evolution Modeling [C5]: For a positive integer t, let $G(t) = (V, E(t))$ be the trust network created within a time period $I(t) = (t0 + (t-1)\Delta t, t0 + t\Delta t]$, where V is the set of nodes that correspond to the individual users in the site at time t0, $E(t)$ ($\subset V \times V$) is the set of trust links created within time-period $I(t)$, and Δt is a positive real number specified in advance. We suppose that there are no self-links and multiple-links. Note that $G0(t) = (V, E(1) \cup \ldots \cup E(t))$ is the trust network for the user set V at time $t0 + t\Delta t$, and $E(s) \cap E(s') = \emptyset$ if $s \neq s'$. We consider predicting the set $E(t+1)$ of trust links created within the next time-period $I(t+1)$. We define a subset $C(t+1)$ of $V \times V$ by $C(t+1) = (V \times V) - \{(v, v) \mid v \in V\} - (E(1) \cup \ldots \cup E(t))$. Then, it is easily seen that $E(t+1) \subset C(t+1)$ and $(E(1) \cup \ldots \cup E(t)) \cap C(t+1) = \emptyset$. Thus, we refer to $C(t+1)$ as the set of candidate trust-links in time-period $I(t+1)$. For any $(u, v) \in C(t+1)$, we investigate whether or not a trust link will be created from node u to node v in the next time-period $I(t+1)$.

We assume that K activities are provided in the site. For any $u \in V$ and positive integer t, let $\mathbf{A}(t; u) = (A(t,1; u), \ldots, A(t,K; u))$ denote the activity vector of node u within time-period $I(t)$, where for each k, $A(t,k; u) = 1$ if user u selected and performed activity k within time-period $I(t)$, and $A(t,k; u) = 0$ otherwise. In this modeling, we aim to investigate the roles of mediators for creating trust links. Thus, we focus on the subset $C_*(t+1)$ of $C(t+1)$ that consists of candidate trust-link $(u, v) \in C(t+1)$ having a mediator $w \in V$ in time-period $I(t+1)$, and for any $(u, v) \in C_*(t+1)$, we consider modeling the probability $P(t+1; u, v)$ that a trust link is created from node u to node v in time-period $I(t+1)$, i.e., $(u, v) \in E(t+1)$. Here, node w is referred to as a mediator from node u to node v in time-period $I(t+1)$ when there exist both a trust link between u and w and a trust link between v and w that are created in time-period It. A mediator w from u to v is classified into four types: w is called type 1 if $(u, w), (w, v) \in E(t)$, w is called type 2 if $(u, w), (v, w) \in E(t)$, w is called type 3 if $(w, u), (w, v) \in E(t)$, and w is called type 4 if $(w, u), (v, w) \in$

$E(t)$. In order to analyze the effects of activities for trust link creation, we aim to investigate the roles of mediators with respect to activities. Thus, we also consider incorporating activity information to model the probability $P(t+1; u, v)$ for any $(u, v) \in C_*(t+1)$.

For any $(u, v) \in C_*(t+1)$, we consider modeling the probability $P(t+1; u, v)$ that trust link (u, v) is created in time-period $I(t+1)$, i.e., $(u, v) \in E(t+1)$. Note that by definition, there exists at least one mediator from node u to node v in $I(t+1)$. In order to analyze the effect of activities for creating a trust link in terms of mediators, we propose two natural stochastic models of $P(t+1; u, v)$. The first model only uses mediator information, and the second model enhances the first model by adding mediator-activity effects for trust-network evolution.

First, we define the A-ME Model. It is conceivable that the presence of mediators affects the creation of trust links. Moreover, we can speculate that the influence strength of a mediator depends on its type. Therefore, in order to analyze the effects of mediators for creating trust links in terms of mediator types, we propose modeling the probability $P(t+1; u, v)$ for any $(u, v) \in C_*(t+1)$ by using a logistic regression model: $P(t+1; u, v) = 1/(1 + \exp(\boldsymbol{\varphi}^{T} \mathbf{y}(t; u, v)))$ where $\boldsymbol{\varphi}$ is a parameter vector, $\boldsymbol{\varphi}^{T} = (\varphi 0, \varphi 1, \varphi 2, \varphi 3, \varphi 4)$, $\mathbf{y}(t; u, v)$ is a feature vector of (u, v) at time $t0 + t\Delta t$, $\mathbf{y}(t; u, v) = (1, y(t,1; u, v), y(t,2; u, v), y(t,3; u, v), y(t,4; u, v))$. Here, each $y(t,i; u, v)$ is the number of type i mediators from u to v in time-period $I(t+1)$. We refer to this stochastic model to simulate the dynamics of creating a trust link as the A-ME model.

Next, we define the A-MAE Model. It is also conceivable that the influence degree of a mediator depends on activity. For $(u, v) \in C_*(t+1)$, let us consider mediators w_k and w_h from node u to node v in time-period $I(t+1)$ such that in time-period $I(t)$, u, v and w_k performed the same activity k, and u, v and w_h did the same activity ℓ, that is, $A(t, k; u) = A(t, k; v) = A(t, k; w_k) = 1$, $A(t, h; u) = A(t, h; v) = A(t, h; w_h) = 1$, where $k \neq h$. Then, for creating a trust link from u to v, the influence that w_k and w_h exert can be different. In order to analyze the effects of activities in terms of mediators, we propose modeling the probability $P(t+1; u, v)$ for any $(u, v) \in C_*(t+1)$ by combining co-occurrence information with respect to activities with the A-ME model: $P(t+1; u, v) = \sum_{k\in\{1,\ldots,K\}} \lambda k/(1 + \exp(\boldsymbol{\theta} k^{T} \mathbf{x}(t,k; u, v)))$, where λk is a parameter vector, $\lambda k = (\lambda 1, \ldots, \lambda K)$; $\lambda 1 + \ldots + \lambda K = 1$, $\lambda k > 0$ $(k = 1, \ldots, K)$, each $\boldsymbol{\theta} k$ is a parameter vector with respect to activity k, $\boldsymbol{\theta} k = (\theta k\_0, \theta k\_1, \theta k\_2, \theta k\_3, \theta k\_4)$, and each $\mathbf{x}(t,k; u, v)$ is a feature vector of (u, v) with respect to activity k at time $t0 + t\Delta t$, $\mathbf{x}(t,k; u, v) = (1, x(t,k\_1; u, v), x(t,k\_2; u, v), x(t,k\_3; u, v), x(t,k\_4; u, v))$. Here, each $x(t,k\_i; u, v)$ is the number of type i mediators w from u to v in time-period $I(t+1)$ such that u, v and w performed activity k in It, that is, $x(t,k\_i; u, v) \geq 0$. In particular, we assume that for any $(u, v) \in C_*(t+1)$, there exist a mediator w′ in time-period $I(t+1)$ and an activity k′ such that nodes u, v and w′ performed activity k′ in $I(t)$, that is, $x(t,k\_i; u, v) > 0$, where w′ is of type i′. We refer to this stochastic model to simulate the dynamics of creating a trust link as the A-MAE model.

In our experiments on trust evolution modeling, we collected real data for a trust network and a set of user activities from Epinions, which is a social media site of product reviews and consumer reports. In Epinions, a user u can create a trust link to another user v by registering v as a trust user. We examined the evolution of the trust network constructed from trust links among users. Also, in Epinions, a user can post a review and give a rating for a product in a given set of products, where those products are classified into K categories. We say that user u performed activity k when u posted a review or gave a rating for some product of category k. By the breadth-first search, we traced in the trust links from a user who was featured as the most popular user in October 2012 until no new users appeared, and collected both a set of trust links and a set of product reviews and ratings. The collected data contains 27, 873 users, 218, 686 trust links, and 809, 521 reviews and 14, 105, 311 ratings for 268, 897 products, where the number of categories was K = 19. On the basis of stability consideration, we exploited only the data generated in 2010, and constructed a dataset from those users that had trust-links and produced activities in 2010. We refer to this dataset as the Epinions data, where the number of users was 749.

User Activity Modeling [C1]: We consider distinguishing a concept of fields which users prefer and a concept of fields for which users gain trust. The former fields are referred to as P-fields, and the latter fields are referred to as T-fields. Unlike hTrust and JLCMF, the proposed NMF model employs two latent spaces. One corresponds to the space of P-fields (called the PF-space), and the other corresponds to the space of T-fields (called the TF-space). Thus, P-field and T-field are also referred to as latent P-factor and latent T-factor, respectively. Let K be the dimension of the PF-space and L the dimension of the TF-space. The proposed NMF model introduces a non-negative N × K matrix **U** = ($U_{i,k}$), a non-negative N × L matrix **W** = ($W_{i,j}$), and a non-negative K×L matrix **H** = ($H_{k,j}$), where $U_{i,k}$ represents the strength of user vi for latent P-factor k, Wi_k represents the strength of user vi for latent T-factor k, and $H_{k,j}$ represents the relationship strength from latent Pfactor k to latent T-factor j for creating trust-links. We consider minimizing the function F (**U**,**W**, **H**) of **U**, **W** and **H**

In our experiments on user activity modeling, we used the datasets for trust networks of @cosme and Epinions explained above. Here, the @Cosme network has 45, 024 nodes and 351, 299 links. For each dataset, we constructed four datasets D1, D2, D3 and D4, by setting the prediction period from January to March for D1, April to June for D2, July to September for D3, and October to December for D4, respectively.

Results and Discussion: Describe significant experimental and/or theoretical research advances or findings and their significance to the field and what work may be performed in the future as a follow on project. Fellow researchers will be interested to know what impact this research has on your particular field of science.

Trust Discount Modeling: We tested the MTDF models with the exponential and power-low discount factors, and evaluated which model is better for the MovieLens dataset. To do this, we computed the log-likelihood ratio statistic of each model against the basic multinomial model for each movie. As the results, we observed a positive correlation, but cannot see a big difference between them, meaning that both decays are equally good and acceptable. Thus, we focused on the rankings based on the z-score derived from the MTDF model with the exponential discount factor. Compared to the rankings of the conventional methods like the average review score over at least 10 posts as shown in Table 1, it is remarkable that the relatively new movies rank in the top-5 thanks to the trust discount factor of the MTDF model that degrades the effects of old reviews, while keeping their average scores comparable with those from the conventional methods as shown in Table 2. Indeed, the ranking of the first-ranked movie is thought reasonable as it is such an acclaimed movie that it won the Academy Awards. On the other hand, the second-ranked movie is also highly ranked by the conventional methods and it is relatively old. This implies that this movie maintains high ratings even in the recent period, and thus it has a high trust level. In summary, the proposed ranking method is useful and derives more reasonable and trustable rankings by threshold. We believe that our MTDF model will play an important role not only for ranking tasks, but also for other tasks such as predicting evolution of social networks.

Table 1: Top 5 movies in the average review score over at least 10 posts

Ranking	Title (year of release)	Avg. score	# of posts
1	The Shawshank Redemption (1994)	4.46	31,126

2	The Godfather (1972)	4.42	19,814
3	The Usual Suspects (1995)	4.37	24,037
4	Schindler's List (1993)	4.36	25,777
5	Sunset Blvd. (1950)	4.32	3,255

Table 2: Top 5 movies in the z-score derived from the MTDF model

Ranking	Title (year of release)	z-score	Avg. score	# of posts
1	The Shawshank Redemption (1994)	157.19	4.46	31,126
2	Schindler's List (1993)	128.85	4.36	25,777
3	The Usual Suspects (1995)	124.96	4.37	24,037
4	The Godfather (1972)	119.82	4.42	19,814
5	The Silence of the Lambs (1991)	119.70	4.20	33,668

User Conformity Modeling: We addressed the problem of quantitatively assessing the conformity of a user in the context of rating items, and proposed an efficient algorithm that learns the conformity metric of each user from observed review scores. The idea behind is that a user often rates an item taking into account not only her own opinion but also scores already given to the item by other users, and the reliability of scores depend on who rated them. We modeled this rating process as a stochastic decision making process and used a modified Voter model. As shown in Figure 1, the proposed method can efficiently learn the conformity metrics based on an iterative algorithm within a few tens of iterations. Its generalization capability is insensitive to the value of the regularization factor. Empirical evaluation on the two real world review datasets uncovered some interesting findings about the conformity metrics learned by the proposed algorithm. As shown in Figure 2, the majority of people have an average conformity metric with adequate regularization factors, i.e., 1.0 and only a limited fraction of people have high or low conformity metrics, who are worth paying attention to. Conformity metric can be a good indicator to identify those who satisfy the following three basic properties simultaneously that are considered natural for a user to be of high conformity, i.e., 1) a multitude of rated items, 2) a multitude of followers, and 3) a high rating similarity between her own scores and her follower's. None of them can be a good indicator alone. We further found that users having a high PageRank score or a high HITS score tend to rate a large number of items and have a large number of followers, satisfying the above two properties, but their rating similarity is not as large as that of those who have high conformity metrics or those who rate a large number of items.

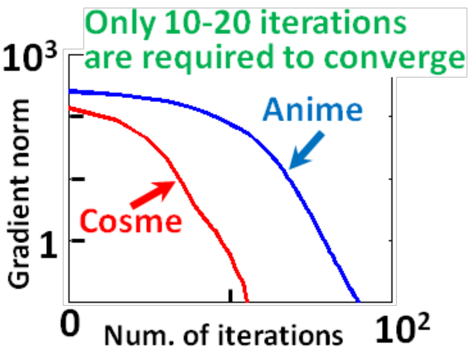

Figure 1: Efficiency of learning algorithm.

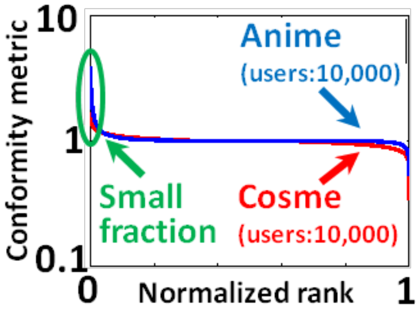

Figure 2: Distribution of conformity metric.

Trust Evolution Modeling: We addressed the problem of modeling the evolution of a trust network in a social media site. In particular, we focused on investigating the roles of mediators for trust-link creation. To this end, we proposed two stochastic models for simulating the dynamics of creating a trust link under the presence of mediators, the A-ME and A-MAE models, where the A-ME model aims to examine mediator effects for trust-network evolution in terms of mediator types, and the A-MAE model enhances the A-ME model to analyze mediator-activity effects for trust-network evolution. For these proposed models, we presented an efficient method of estimating the values of parameters from an observed sequence of trust links and user activities. Using real data from Epinions, we experimentally evaluated the A-ME and A-MAE models for predicting trust links in the near future under the presence of mediators. First, by comparing the A-ME model and random guessing, we demonstrated that incorporating mediator-type information has a positive effect for predicting trust-links. Next, we showed that the A-MAE model significantly outperforms the A-ME model, and demonstrated the effectiveness of mediator-activity information for trust-network evolution, in two cases with mediator weights and without them as shown in Figure 3. We also showed that different mediator-activities differently affect trust-link creation, and different mediator types differently affect trust-link creation. Moreover, by using the A-ME and A-MAE models, we found several characteristic properties of trust-link creation probability in the Epinions data in terms of mediator-activities and mediator-types.

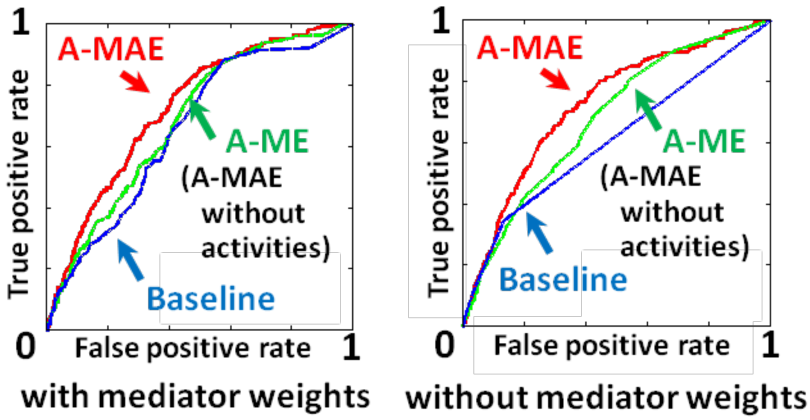

Figure 3: Effectiveness of mediator-activity information.

User Activity Modeling: We addressed the problem of modeling activities among users for an item-review site. To this end, we proposed a new NMF method that incorporates people's evaluation of users' activities as well as trust-links and users' activities themselves. In our experiments using two real world item-review sites, @cosme and Epinions, we statistically analyzed the datasets, and in particular confirmed that the number of appreciation messages received correlates with the number of trust-links received, suggesting that incorporating the activity-evaluation information can be a promising approach. Next, we demonstrated that in terms of the area under the ROC curve (AUC), the proposed method outperforms hTrust and its variants JLCMF and JLCMF2 in a trust-link prediction problem as show in Figures 4 and 5, which correspond to the results for the @cosme and Epinions datasets, respectively. Further, we applied the proposed method to an analysis of users' behavior in an item-review site, and found several characteristic properties for @cosme and Epinions from the perspective of trust-link creation.

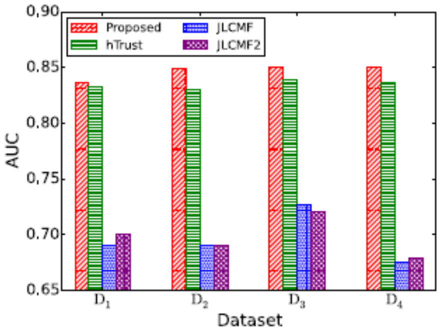

Figure 4: @cosme dataset.

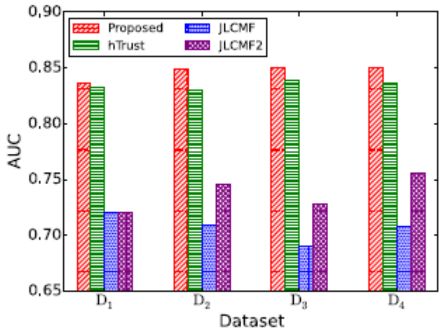

Figure 5: Epinions dataset

Note: Final report is for the entire project period, not just for the last one year. Section structure need not be the same as above. You can structure the report in the same way as you normally write a journal paper.

List of Publications and Significant Collaborations that resulted from your AOARD supported project: In standard format showing authors, title, journal, issue, pages, and date, for each category list the following:

a) papers published in peer-reviewed journals,

J1. Masahiro Kimura, Kazumi Saito, Kouzou Ohara, and Hiroshi Motoda, "Speeding-up Node Influence Computation for Huge Social Networks," International Journal of Data Science and Analytics, (Online First: DOI 10.1007/s41060-015-0001-y), 2016.

J2. Kazumi Saito, Masahiro Kimura, Kouzou Ohara, and Hiroshi Motoda, "A new centrality measure of node importance for information diffusion over social network," Information Sciences, Vol.329, pp.985--1000, Feb. 2016.

J3. Kazumi Saito, Kouzou Ohara, Masahiro Kimura, and Hiroshi Motoda, "Change Point Detection for Burst Analysis from an Observed Information Diffusion Sequence of Tweets," Journal of Intelligent Information Systems (JIIS), Vol.44, Iss.2, pp. 243-269, 2015.

b) papers published in peer-reviewed conference proceedings,

C1. Kanji Matsutani, Masahito Kumano, Masahiro Kimura, Kazumi Saito, Kouzou Ohara and Hiroshi Motoda, "Combining Activity-evaluation Information with NMF for Trust-link Prediction in Social Media," Proc. of the 2015 IEEE International Conference on Big Data (BigData2015), pp.2101--2110, 2015.

C2. Kouzou Ohara, Kazumi Saito, Masahiro Kimura and Hiroshi Motoda, "Change Point Detection for Information Diffusion Tree," Proc. of the Eighteenth International Conference on Discovery Science (DS2015), pp.161--169, 2015.

C3. Kouzou Ohara, Kazumi Saito, Masahiro Kimura and Hiroshi Motoda, "Resampling-based Gap Analysis for Detecting Nodes with High Centrality on Large Social Network," Proc. of the Nineteenth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2015) , pp.135--147, 2015.

C4. Kazumi Saito, Masahiro Kimura, Kouzou Ohara and Hiroshi Motoda, "Efficient Learning of User Conformity on Review Score," Proc. of the 2015 International Conference on Social Computing, Behavioral Modeling, and Prediction (SBP2015), pp. 182–192, 2015.

C5. Keito Hatta, Masahito Kumano, Masahiro Kimura, Kazumi Saito, Kouzou Ohara and Hiroshi Motoda, "Analyzing Mediator-Activity Effects for Trust-Network Evolution in Social Media, " Proc. of the 13th Pacific Rim International Conference on Artificial Intelligence (PRICAI2014), pp.297-308, 2014.

C6. Yuki Yamagishi, Seiya Okubo, Kazumi Saito, Kouzou Ohara, Masahiro Kimura, and Hiroshi Motoda, "A Method to Divide Stream Data of Scores over Review Sites, " Proc. of the 13th Pacific Rim International Conference on Artificial Intelligence (PRICAI2014),

pp.913-919, 2014.

C7. Kouzou Ohara, Kazumi Saito, Masahiro Kimura and Hiroshi Motoda, "Resampling-based Framework for Estimating Node Centrality of Large Social Network," Proc. of the Seventeenth International Conference on Discovery Science (DS2014), pp.228-239, 2014.

C8. Masahiro Kimura, Kazumi Saito, Kouzou Ohara and Hiroshi Motoda, "Efficient Analysis of Node Influence Based on SIR Model over Huge Complex Networks, " Proc. of the 2014 International Conference on Data Science and Advanced Analytics (DSAA2014), pp.S7-1:1-S7-1:7, 2014.

C9. Kazumi Saito, Masahiro Kimura, Kouzou Ohara and Hiroshi Motoda, " New Approach for Item Ranking Based on Review Scores Reflecting Temporal Trust Factor," Proc. of the 2014 International Conference on Social Computing, Behavioral Modeling, and Prediction (SBP2014), pp. 157--164, 2014.

C10. Kouzou Ohara, Kazumi Saito, Masahiro Kimura and Hiroshi Motoda, "Predictive Simulation Framework of Stochastic Diffusion Model for Identifying Top-K Influential Nodes," Proc. of the Fifth Asian Conference on Machine Learning (ACML2013), pp.149--164, 2013.